

# Gestión de Retención de Empleados

Jorge Luis Aquino Olmos

2024-03-12

## Modelos de Regresión

Los modelos de regresión son una clase de técnicas estadísticas utilizadas para predecir el valor de una variable dependiente basándose en una o varias variables independientes. Las variables independientes, también llamadas predictoras o variables explicativas, son aquellas que se usan para realizar la predicción.

Existen varios tipos de modelos de regresión, cada uno diseñado para diferentes tipos de datos y situaciones, algunos de los tipos más comunes de modelos de regresión:

1. **Regresión Lineal simple:** Es el tipo más básico de modelo de regresión. Se utiliza cuando hay una relación lineal entre una variable independiente y una variable dependiente. La ecuación de regresión lineal simple tiene la forma:

$$y = \beta_0 + \beta_1 x + \epsilon$$

donde  $y$  es la variable dependiente,  $x$  es la variable independiente  $\beta_0$  es el intercepto,  $\beta_1$  es la pendiente y  $\epsilon$  es el término de error.

2. **Regresión lineal múltiple:** Este modelo se utiliza cuando hay más de una variable independiente que influye en la variable dependiente. La ecuación de regresión lineal múltiple tiene la forma

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon$$

donde  $\beta_0, \beta_1, \dots, \beta_n$  son los coeficientes que se estiman a partir de los datos  $x_1, x_2, \dots, x_n$  y  $\epsilon$  es el término de error.

3. **Regresión polinomial:** Este tipo de regresión se utiliza cuando la relación entre la variable independiente y la variable dependiente no es lineal. Se ajusta una curva polinomial a los datos, lo que permite capturar relaciones más complejas. Por ejemplo, una regresión polinomial de segundo grado tiene la forma

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$$

4. **Regresión Logística:** La regresión logística, también conocida como “modelo logit”, es una técnica estadística utilizada para predecir variables categóricas a partir de variables predictoras. A diferencia de la regresión lineal, que se emplea para predecir valores continuos, la regresión logística se aplica cuando la variable dependiente es finita o categórica. Esta variable puede ser binaria, como sí/no o 1/0, lo que se conoce como regresión binaria, o puede tener múltiples categorías, como A, B, C o D, lo que se conoce como regresión multinomial.

## Estimación de Coeficientes mediante Mínimos Cuadrados Ordinarios

Los Mínimos Cuadrados Ordinarios (OLS, por sus siglas en inglés) es un método comúnmente utilizado para estimar los parámetros desconocidos en modelos de regresión. El objetivo es encontrar los coeficientes que

minimizan la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos por el modelo.

Para el caso de la regresión lineal simple, la estimación de los coeficientes se realiza de la siguiente manera:

Dado un conjunto de datos  $(x_i, y_i)$  para  $i = 1, 2, \dots, n$ , donde  $x_i$  es la variable independiente y  $y_i$  es la variable dependiente, la estimación de los coeficientes  $\beta_0$  y  $\beta_1$  se obtiene minimizando la función de pérdida, que en este caso es la suma de los cuadrados de los residuos (errores), denotada por  $SSR$ :

$$SSR = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \quad (1)$$

Para encontrar los valores de  $\beta_0$  y  $\beta_1$  que minimizan  $SSR$ , se derivan parcialmente  $SSR$  con respecto a  $\beta_0$  y  $\beta_1$ , se igualan a cero y se resuelven las ecuaciones resultantes. Esto conduce a las llamadas ecuaciones normales:

$$\frac{\partial SSR}{\partial \beta_0} = -2 \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i)) = 0 \quad (2)$$

$$\frac{\partial SSR}{\partial \beta_1} = -2 \sum_{i=1}^n x_i (y_i - (\beta_0 + \beta_1 x_i)) = 0 \quad (3)$$

Resolviendo estas ecuaciones, se obtienen las estimaciones de los coeficientes  $\beta_0$  y  $\beta_1$ :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (4)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (5)$$

donde  $\bar{x}$  y  $\bar{y}$  son las medias de las variables independiente y dependiente, respectivamente.

Para la regresión lineal múltiple y otros tipos de regresión, las ecuaciones normales se generalizan en forma matricial:

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \quad (6)$$

donde  $\mathbf{X}$  es una matriz de diseño que contiene las variables independientes,  $\mathbf{y}$  es un vector de la variable dependiente,  $\hat{\beta}$  son los coeficientes estimados.

## Modelización de Retención de Empleados y Ganancias de Tiendas en Store24

El objetivo de este estudio es comprender cómo la retención de empleados en las tiendas de Store24 influye en sus ganancias. Utilizando datos del año fiscal 2000, exploraremos diversas variables relacionadas con la gestión de recursos humanos, características de ubicación de la tienda y datos demográficos para determinar su impacto en la utilidad de las tiendas.

Nuestra variable objetivo será la “Utilidad del año fiscal 2000 antes de asignación de gastos indirectos corporativos, alquiler y depreciación”, que refleja las ganancias de cada tienda. Consideraremos una serie de variables predictoras, como la antigüedad promedio de los gerentes y el personal, la competencia en la ubicación de la tienda, la densidad de población circundante y características de la tienda, como su horario de apertura y su ubicación residencial o industrial.

A través de técnicas de regresión, exploraremos la relación entre estas variables y las ganancias de las tiendas, con el objetivo de identificar áreas de oportunidad para mejorar la retención de empleados y, en última instancia, aumentar las ganancias de Store24.

## Variables que haran parte de estudio.

Variable	Descripción
Sales	Ventas del año fiscal 2000
Profit	Utilidad del año fiscal 2000 antes de asignación de gastos indirectos corporativos, alquiler y depreciación
MTenure	Antigüedad promedio en el puesto del gerente durante el año fiscal 2000 donde la tenencia se define como el número de meses de experiencia en Store24
CTenure	Antigüedad promedio en el puesto del personal durante el año fiscal 2000 donde la tenencia se define como el número de meses de experiencia en Store24
Comp	Número de competidores por 10,000 personas dentro de un radio de 1/2 milla
Pop	Población dentro de un radio de 1/2 milla
Visibility	Calificación de 5 puntos en visibilidad del frente de la tienda, siendo 5 la más alta
PedCount	Calificación de 5 puntos sobre el volumen de tráfico de peatones, siendo 5 el más alto
Hours24	Indicador de si la tienda abre o no 24 horas
Res	Indicador de ubicado en zona residencial vs. industrial
CrewSkill	Habilidad del equipo
MgrSkill	Habilidad de gestión
ServQual	Medición de calidad de servicio

Como podemos observar el conjunto de datos que servirán para realizar nuestro modelo contienen información financieras, otras que describen la gestión del recurso humano y variables de ubicación de las tiendas.

## Lectura de los datos

En nuestro conjunto de datos, contamos con varias variables que representan la valoración de cualidades específicas en un contexto particular. Estas variables, denominadas Visibility, PedCount, Res y Hours24, están diseñadas para capturar la evaluación de ciertas características relevantes para el desempeño y la operación de las tiendas. Cada una de estas variables de valoración de cualidades tiene un conjunto discreto de valores que representan diferentes niveles de la cualidad evaluada. Al ser variables categóricas ordinales, estas valoraciones no solo nos brindan una comprensión de la situación actual de las tiendas en términos de las cualidades evaluadas, sino que también nos permiten identificar tendencias, patrones y áreas de mejora en nuestra operación y gestión.

A continuación, se presenta un resumen estadístico de las variables numéricas en nuestro conjunto de datos. Este resumen proporciona una visión general de la distribución y variabilidad de cada variable, lo que nos ayuda a comprender mejor la naturaleza de nuestros datos y a identificar posibles patrones o tendencias.

##	vars	n	mean	sd	median	trimmed	mad
## Sales	1	75	1205413.12	304531.31	1127332.00	1182031.25	288422.04
## Profit	2	75	276313.61	89404.08	265014.00	270260.34	90532.00
## MTenure	3	75	45.30	57.67	24.12	33.58	29.67
## CTenure	4	75	13.93	17.70	7.21	10.60	6.14
## Pop	5	75	9825.59	5911.67	8896.00	9366.07	7266.22
## Comp	6	75	3.79	1.31	3.63	3.66	0.82
## CrewSkill	7	75	3.46	0.41	3.50	3.47	0.34
## MgrSkill	8	75	3.64	0.41	3.59	3.62	0.45
## ServQual	9	75	87.15	12.61	89.47	88.62	15.61
##		min	max	range	skew	kurtosis	se
## Sales	699306.00	2113089.00	1413783.00	0.71	-0.09	35164.25	
## Profit	122180.00	518998.00	396818.00	0.62	-0.21	10323.49	
## MTenure	0.00	277.99	277.99	2.01	3.90	6.66	

## CTenure	0.89	114.15	113.26	3.52	15.00	2.04
## Pop	1046.00	26519.00	25473.00	0.62	-0.23	682.62
## Comp	1.65	11.13	9.48	2.48	11.31	0.15
## CrewSkill	2.06	4.64	2.58	-0.43	1.64	0.05
## MgrSkill	2.96	4.62	1.67	0.27	-0.53	0.05
## ServQual	57.90	100.00	42.10	-0.66	-0.72	1.46

Nuestro resumen estadístico incluye las siguientes estadísticas para cada una de las nueve variables numéricas en nuestro conjunto de datos:

#### Medidas de tendencia central:

- **Media:** La media es el promedio de todos los valores en un conjunto de datos. Es una medida de tendencia central que indica el valor “típico” de una variable. Mediana: La mediana es el valor que divide un conjunto de datos en dos mitades iguales. Es una medida de tendencia central que no se ve afectada por valores atípicos. Moda: La moda es el valor que aparece con mayor frecuencia en un conjunto de datos. Es una medida de tendencia central que indica el valor más común. Medidas de dispersión:
- **Desviación estándar:** La desviación estándar es una medida de la variabilidad de los datos alrededor de la media. Un valor alto de desviación estándar indica que los datos están muy dispersos, mientras que un valor bajo indica que los datos están muy concentrados alrededor de la media.
- **Rango:** El rango es la diferencia entre el valor máximo y el valor mínimo en un conjunto de datos. Es una medida simple de la dispersión de los datos.

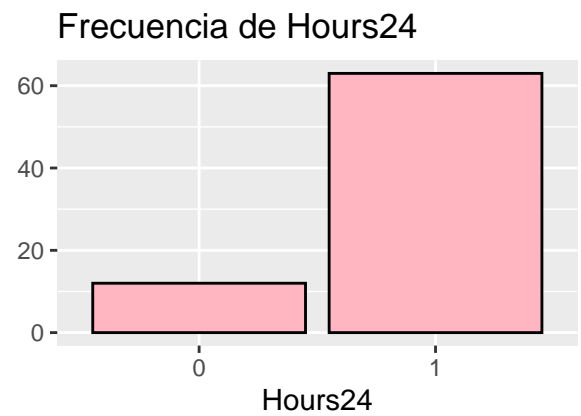
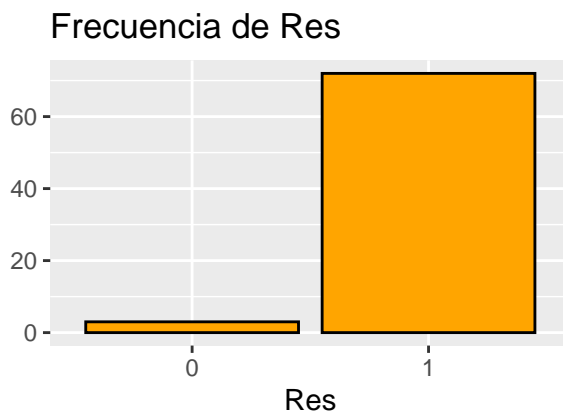
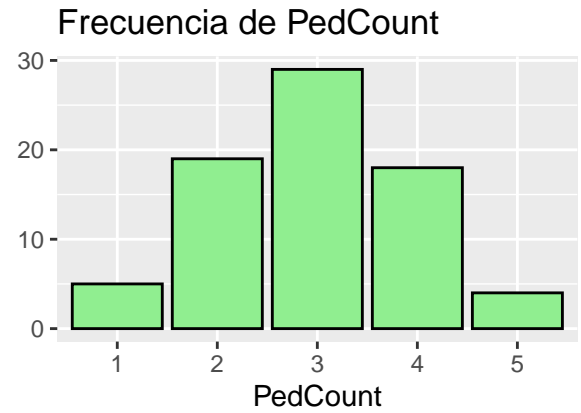
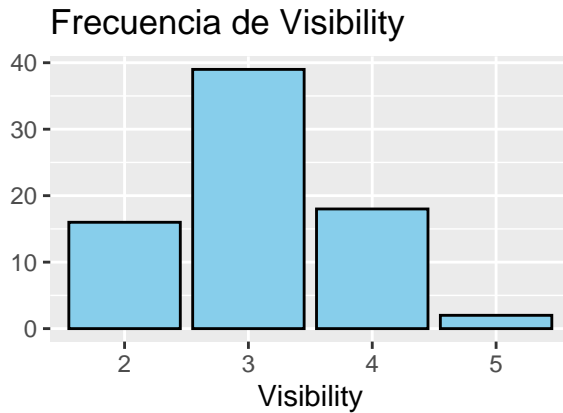
#### Medidas de forma:

- **Asimetría:** La asimetría indica si la distribución de los datos es simétrica o no. Un valor positivo de asimetría indica que la distribución está sesgada hacia la derecha, mientras que un valor negativo indica que la distribución está sesgada hacia la izquierda.
- **Curtosis:** La curtosis indica si la distribución de los datos es más o menos aplanada que una distribución normal. Un valor positivo de curtosis indica una distribución más puntiaguda que la normal, mientras que un valor negativo indica una distribución más aplanada que la normal.

#### Otras medidas:

-**Error estándar:** El error estándar es una medida de la precisión de una estimación estadística. Es una medida de cuánto se espera que la estimación difiera del valor real.

## Descripción de las variables categóricas



### Visibilidad:

- La mayoría de las tiendas tienen una visibilidad alta (4 o 5).
- Hay algunas tiendas con una visibilidad baja (1 o 2).

### Tráfico peatonal:

- La mayoría de las tiendas tienen un tráfico peatonal medio (3).
- Hay algunas tiendas con un tráfico peatonal alto (4 o 5) y bajo (1 o 2).

### Horario de apertura:

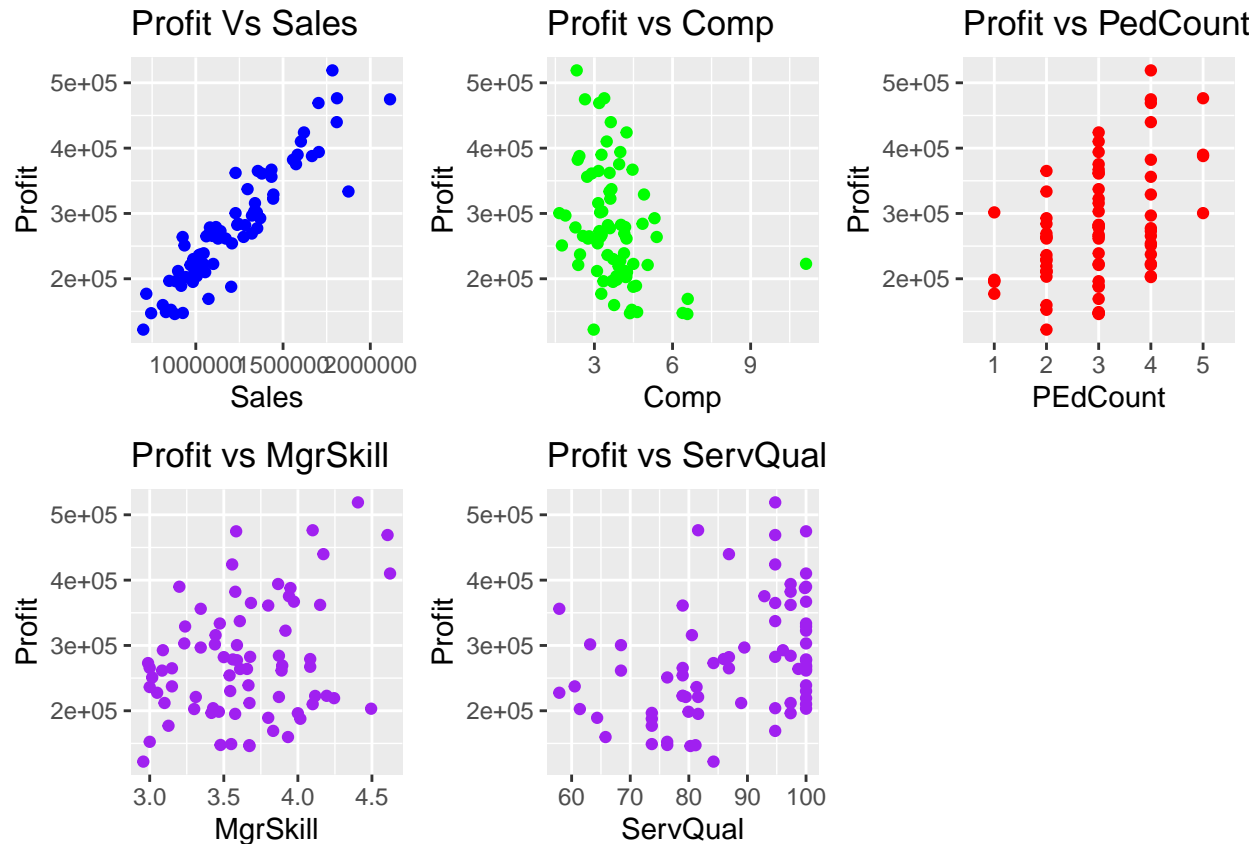
- La mayoría de las tiendas no abren las 24 horas (0).
- Hay algunas tiendas que sí abren las 24 horas (1).

### Ubicación:

- La mayoría de las tiendas están ubicadas en zonas residenciales (1).
- Hay algunas tiendas que están ubicadas en zonas industriales (0).

## Visualización de la relación existe entre las variables

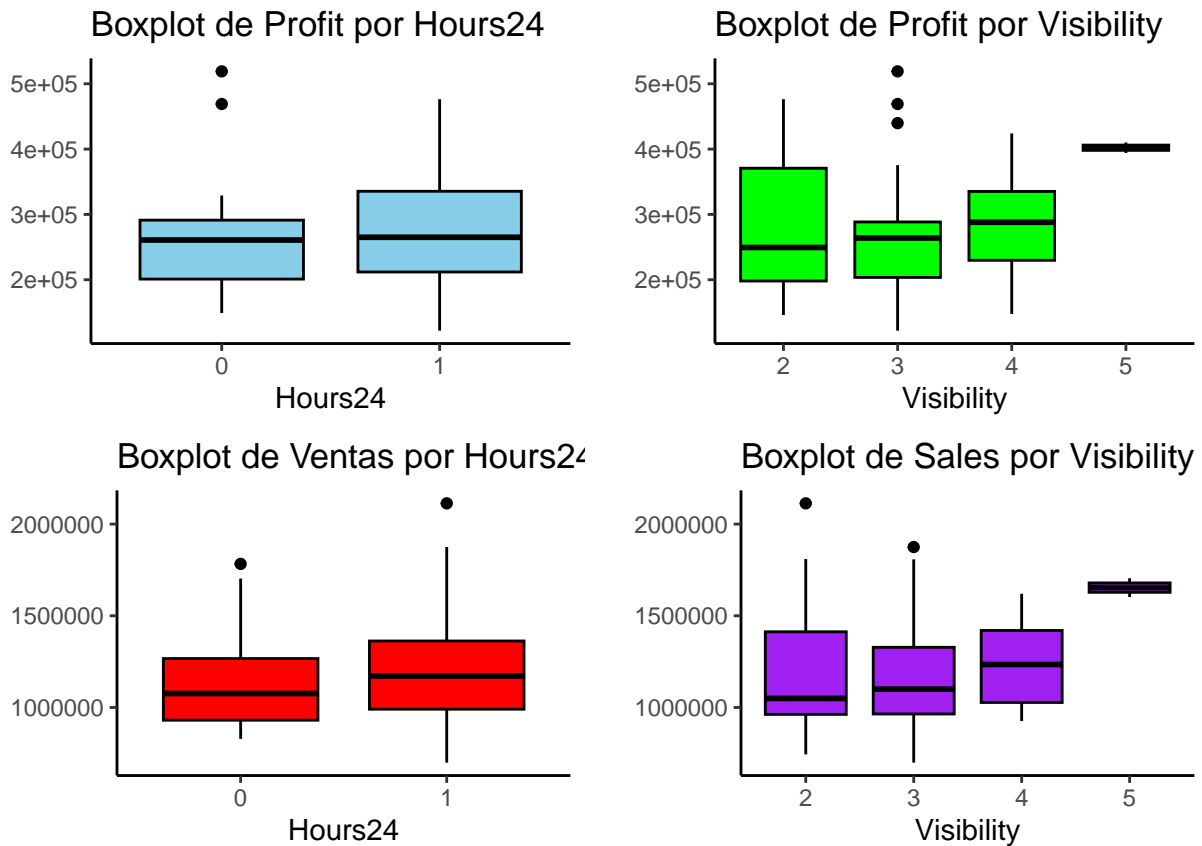
Procederé a realizar algunas visualizaciones con el fin de obtener una comprensión intuitiva de la relación que pueda existir entre el beneficio (Profit) y algunas variables presentes en el conjunto de datos.



Estas visualizaciones muestran cómo algunas variables seleccionadas se relacionan con Profit. Observamos que:

- **Sales vs Profit:** Hay una tendencia clara que un aumento en las ventas está asociado con un aumento en las ganancias.
- **Comp (Competencia) vs Profit:** A medida que el nivel de competencia aumenta, hay una tendencia a disminuir las ganancias, aunque la relación no parece ser tan fuerte.
- **PedCount (Conteo de Peatones) vs Profit:** Existe cierta tendencia positiva, indicando que un mayor tráfico peatonal podría estar relacionado con mayores ganancias.
- **MgrSkill (Habilidad del Gerente) vs Profit:** Se observa una relación positiva moderada, sugiriendo que las habilidades de gestión más altas pueden estar asociadas con mejores ganancias.
- **ServQual (Calidad del Servicio) vs Profit:** También se observa una relación positiva, indicando que una mejor calidad del servicio puede contribuir a mayores ganancias.

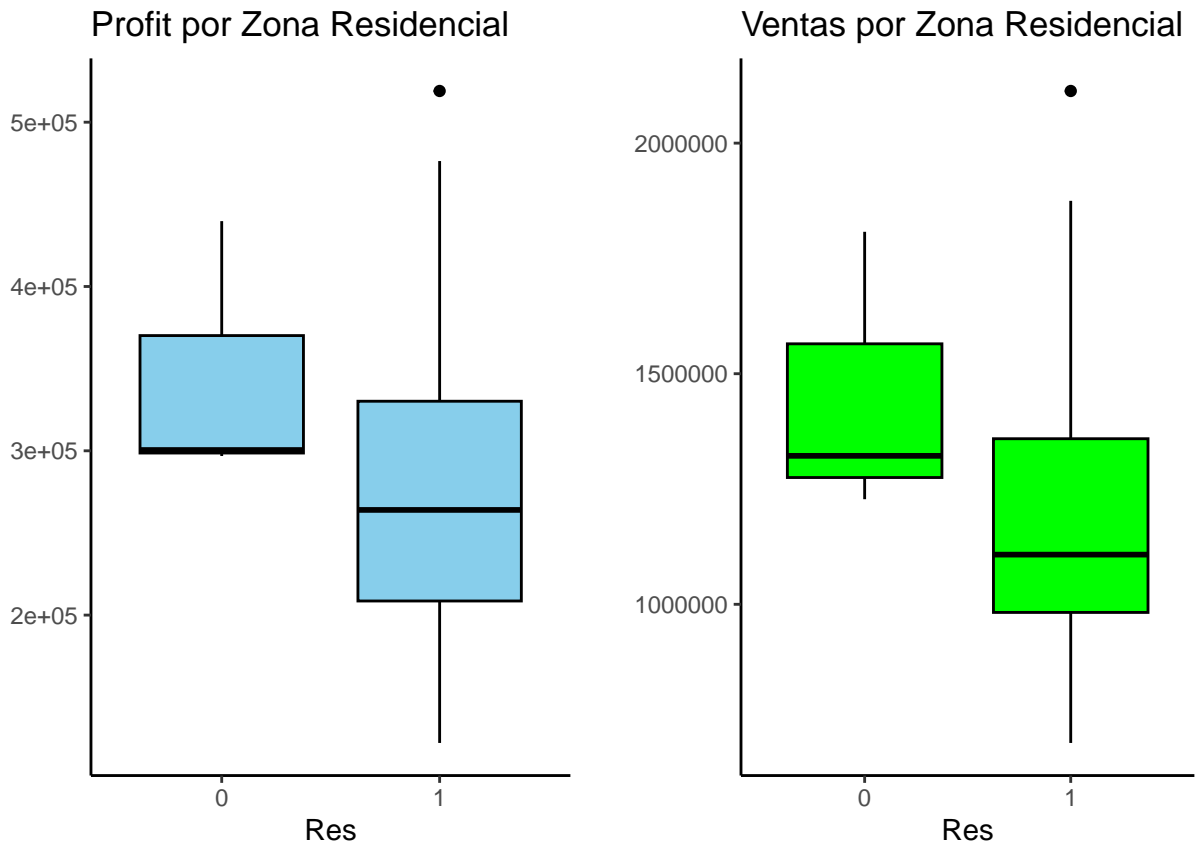
### Boxplot de Profit y Sales para las variables Hours24 y Visibility



Los gráficos de caja y bigotes revelan patrones interesantes en relación con el desempeño de las tiendas. Es notable que las tiendas que operan las 24 horas del día exhiben tanto un beneficio más alto como mayores ventas en comparación con las tiendas que tienen un horario de apertura estándar. Este hallazgo sugiere una correlación entre la disponibilidad continua y el éxito financiero de las tiendas.

Además, al observar la visibilidad de las tiendas, destacamos que aquellas que recibieron una calificación perfecta de 5 muestran un desempeño sobresaliente en términos de ganancias y ventas. Estas tiendas con una visibilidad excepcional experimentan un incremento significativo en sus ingresos en comparación con las tiendas que recibieron calificaciones más bajas. Esto sugiere una relación positiva entre la visibilidad del frente de la tienda y su rendimiento financiero, lo que subraya la importancia de una ubicación estratégica y una presentación visual atractiva para impulsar el éxito comercial.

## Beneficio y Ventas por tipo de zona

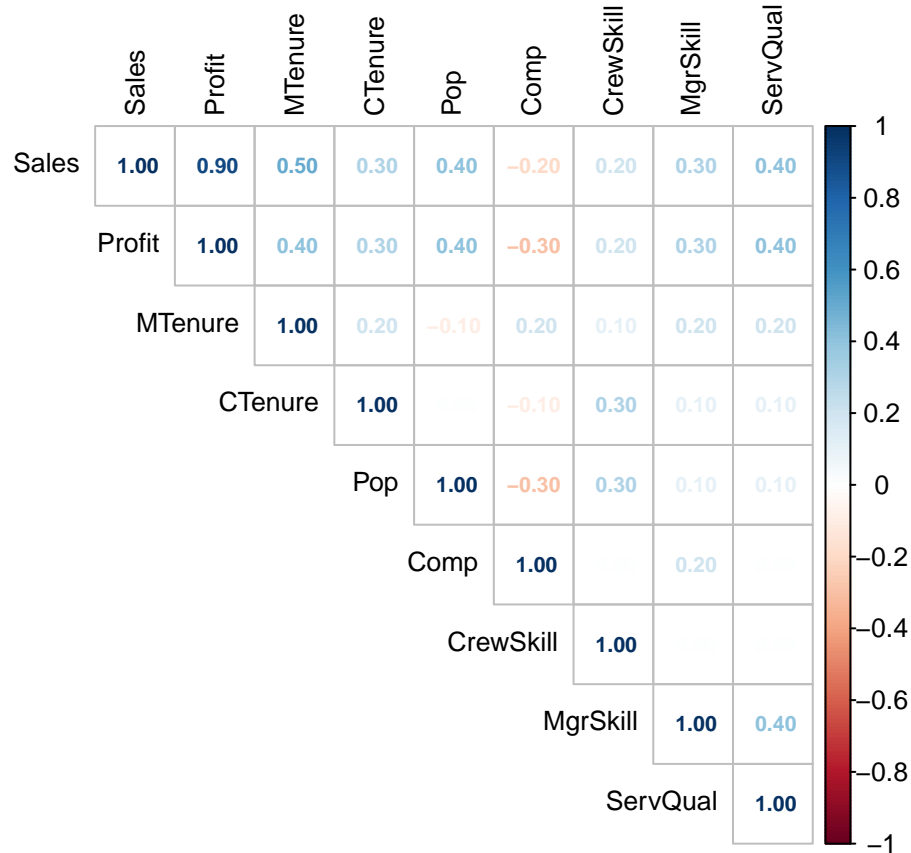


Al analizar los gráficos previos, se evidencia que la variabilidad de las ventas y las ganancias en las tiendas situadas en zonas residenciales es considerablemente menor en comparación con las tiendas ubicadas en áreas industriales. Por otro lado, las tiendas en zonas industriales exhiben una variabilidad más alta, lo que indica una mayor sensibilidad a factores externos o una mayor diversidad en los patrones de consumo en estas áreas.

## Mapa de correlación

Un mapa de correlación es una herramienta visual poderosa que nos permite explorar las relaciones entre variables en un conjunto de datos. Esta representación gráfica nos proporciona una visión general de cómo las diferentes variables están relacionadas entre sí, mostrando la fuerza y la dirección de las asociaciones. Los coeficientes de correlación se representan mediante colores o mediante un código de colores, lo que facilita la identificación de patrones y tendencias en los datos. Este análisis es fundamental en la exploración de datos y en la identificación de posibles relaciones significativas que puedan influir en nuestro análisis o en la toma de decisiones.





- Hay una fuerte correlación positiva entre el beneficio y las ventas (Profit ~ Sales).
- Hay una correlación positiva débil entre el beneficios y la antigüedad en el puesto del gerente (Profit ~ MTenure).
- Hay una correlación positiva débil entre el beneficio y la calidad del servicio (Profit ~ SerQual).
- Hay una correlación negativa débil entre el beneficio y la competencia (Profit ~ Comp).
- Hay una correlación positiva débil entre la antigüedad en el puesto del personal y la habilidad del equipo (CTenure ~ CrewSkill).
- Hay una correlación positiva débil entre la antigüedad en el puesto del gerente y la habilidad en la gestión (MTenure ~ MgrSkill).

## Valores atípicos

Los valores atípicos, también conocidos como valores extremos o anomalías, son observaciones que se desvían significativamente del patrón general del conjunto de datos. En el contexto de la regresión lineal, los valores atípicos pueden tener un impacto considerable en la precisión y la interpretación del modelo. Por lo tanto, es fundamental estudiar y comprender la presencia de valores atípicos en el análisis de regresión, por ello se presentará un cuadro donde se indican cuántos **Outliers** hay presente en cada variable numérica:

Table 2: Conteo de valores atípicos por variable

Variable	Valores_Atípicos
Sales	1
Profit	1
MTenure	9

Variable	Valores_Atípicos
CTenure	4
Pop	0
Comp	4
CrewSkill	4
MgrSkill	0
ServQual	0

## Estandarizacion de los datos

Considerando la presencia de valores atípicos importantes en nuestros datos financieros y su potencial relevancia para nuestro análisis, haré uso de la estandarización como método de preprocesamiento de datos. La estandarización es menos sensible a los valores atípicos en comparación con la normalización, ya que utiliza la media y la desviación estándar de los datos para centrar y escalar las características. Esto significa que los valores atípicos tienen menos influencia en la escala de las características estandarizadas, lo que puede ayudar a mitigar su impacto en nuestro modelo de regresión lineal múltiple. Además, la estandarización facilita una interpretación más clara de los coeficientes del modelo, ya que representan el cambio en la variable de respuesta en términos de desviaciones estándar de las características correspondientes.

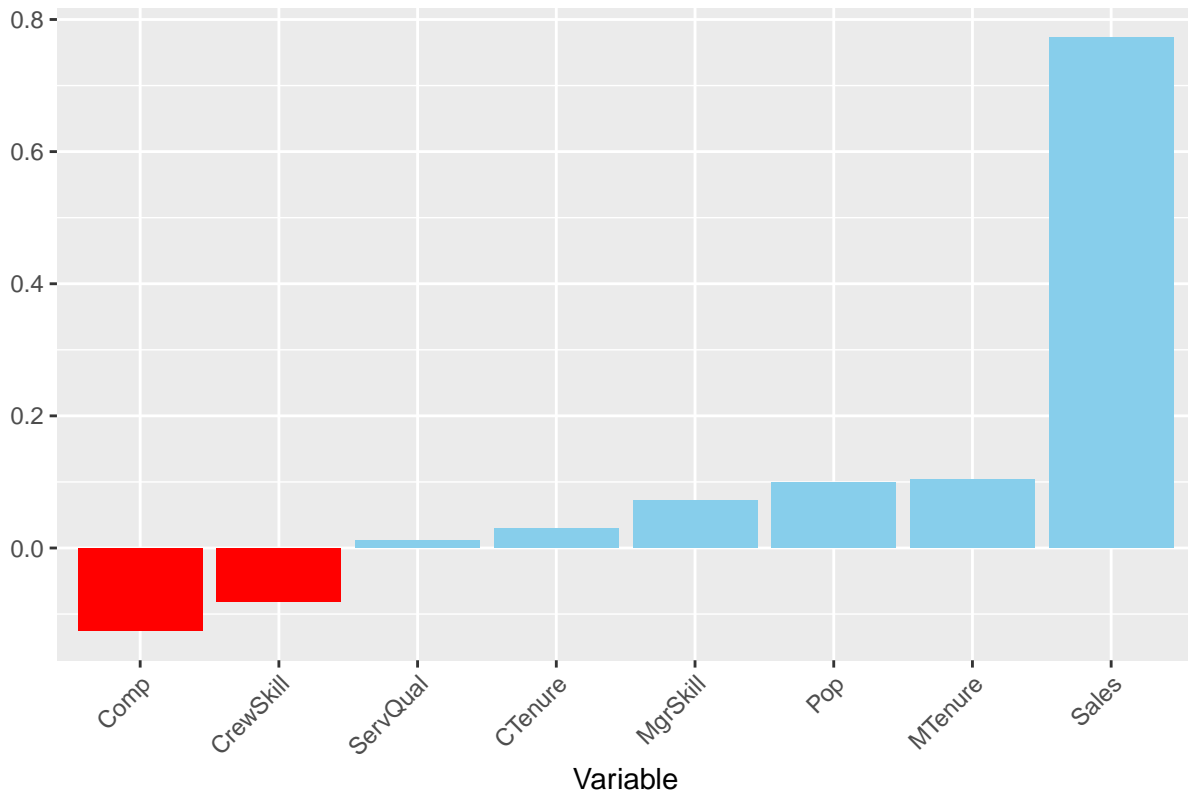
## Creación del modelo

Para construir el modelo de regresión lineal, utilizaré la función `lm()` del paquete `stats` de R. Esta función nos permitirá crear el modelo y obtener una visión detallada de cómo se comporta el mismo. A través de `lm()`, podremos ajustar el modelo a nuestros datos y examinar los coeficientes de regresión, los residuos, así como otras métricas de evaluación que nos ayudarán a comprender la relación entre las variables predictoras y la variable de respuesta.

```
##
##
## | Métrica | Valor |
## |-----:|-----:|
## | RMSE (Error Cuadrático Medio Raíz) | 0.3854082 |
## | R2 (Coeficiente de Determinación) | 0.8840036 |
```

```
##
##
## Table: Coeficientes del Modelo de Regresión
##
## | Variable | Coeficiente |
## |-----:|-----:|
## | (Intercept) | (Intercept) | -0.0115501 |
## | Sales | Sales | 0.7729019 |
## | MTenure | MTenure | 0.1032391 |
## | CTenure | CTenure | 0.0300994 |
## | Pop | Pop | 0.0990065 |
## | Comp | Comp | -0.1250342 |
## | CrewSkill | CrewSkill | -0.0802951 |
## | MgrSkill | MgrSkill | 0.0723050 |
## | ServQual | ServQual | 0.0111320 |
```

## Coeficientes del Modelo de Regresión Lineal



## Resumen del Modelo

##	Variable	Coeficiente	P.value	Significancia
##	(Intercept)	(Intercept)	-0.01155006	8.089129e-01
##	Sales	Sales	0.77290191	8.185012e-16 *****
##	MTenure	MTenure	0.10323910	1.677866e-01
##	CTenure	CTenure	0.03009935	5.331914e-01
##	Pop	Pop	0.09900652	8.362654e-02 *****
##	Comp	Comp	-0.12503416	2.369793e-02 *****
##	CrewSkill	CrewSkill	-0.08029510	2.180742e-01
##	MgrSkill	MgrSkill	0.07230499	1.912647e-01
##	ServQual	ServQual	0.01113196	8.382281e-01

## Interpretación de los resultados

- **RMSE (Error Cuadrático Medio Raíz):** Un valor de RMSE de 0.3854082 indica que, en promedio, las predicciones del modelo están desviadas por aproximadamente 0.3854082 unidades de la variable objetivo. Dado que la variable objetivo oscila entre -1.724011 y 2.71446, un RMSE de esta magnitud podría considerarse aceptable, especialmente si se tiene en cuenta la escala y la variabilidad de la variable objetivo.
- **R<sup>2</sup> (Coeficiente de Determinación):** Un valor de R<sup>2</sup> de 0.8840036 sugiere que el modelo explica aproximadamente el 88.4% de la variabilidad en la variable objetivo. Esto indica que el modelo tiene una capacidad predictiva considerable y puede explicar una cantidad significativa de la variabilidad en los datos observados.

En resumen, los resultados sugieren que el modelo tiene un buen desempeño en la predicción de la variable objetivo y puede ser útil para comprender y predecir el comportamiento de la variable objetivo en función de las variables predictoras incluidas en el modelo.

### Variables significativas

En nuestro análisis, la variable Sales emerge como la más significativa, seguida, aunque no de manera inmediata, por MTenure (Antigüedad en el puesto del gerente), cuyo impacto resulta ser bastante tenue. Por otro lado, la variable Comp (competidores) muestra una relación inversa con las ganancias, indicando que una mayor competencia se correlaciona con menores ganancias. Sin embargo, estas observaciones, aunque reveladoras, no proporcionan una respuesta completa a nuestra pregunta inicial. Es posible que otras variables no consideradas en nuestro análisis puedan desempeñar un papel crucial en las ganancias de las tiendas.

Para profundizar en esta investigación, proponemos varias vías de análisis adicionales:

1. **Análisis de Interacción:** Explorar si la combinación de variables, como antigüedad en el puesto del gerente y la habilidad de gestión, tiene un efecto sinérgico en las ganancias. Este enfoque podría revelar si las tiendas con un gerente con varios años de servicio obtienen un mayor beneficio que aquellas con gerente sin.
2. **Análisis de Subgrupos:** Investigar cómo estas variables afectan las ganancias en diferentes contextos, como ubicaciones específicas o tiendas de diferentes tamaños. Esto podría proporcionar información sobre si la importancia de la habilidad del gerente varía según el entorno.
3. **Importancia Relativa:** Emplear técnicas estadísticas para evaluar la importancia relativa de las variables en la predicción de ganancias. Este análisis arrojaría luz sobre cuánto contribuye cada factor al éxito financiero de las tiendas.

Al profundizar de esta manera, podemos obtener una comprensión más completa de los factores que influyen en las ganancias de las tiendas y desarrollar estrategias más efectivas para optimizar el rendimiento financiero.

### Añadir un término de interacción al DataFrame

En el análisis de regresión lineal, la construcción de modelos precisos y representativos es esencial para comprender las relaciones entre las variables predictoras y la variable de respuesta. Una técnica poderosa para mejorar la capacidad predictiva de un modelo es la creación de variables de interacción.

Las variables de interacción son productos entre dos o más variables predictoras y pueden capturar efectos conjuntos que no se pueden capturar considerando las variables por separado. Esta técnica es particularmente útil cuando se sospecha que las relaciones entre las variables predictoras y la variable de respuesta son no lineales o cuando ciertas variables tienen un efecto modificador en otras.

Para nuestro estudio, nos centramos en el beneficio (Profit) de una empresa y cómo está influenciado por diversas variables, incluyendo la antigüedad en el puesto (MTenure y CTenure) y las habilidades del personal y los gerentes (CrewSkill y MgrSkill). Es crucial entender cómo la experiencia acumulada del personal y los gerentes interactúa con sus habilidades respectivas para influir en el rendimiento de la empresa.

Por tanto, hemos creado dos nuevas variables de interacción:

1. **Interacción entre CTenure y CrewSkill:** Esta variable captura cómo la antigüedad en el puesto del personal interactúa con sus habilidades. Es interesante observar cómo la experiencia acumulada del personal se relaciona con sus habilidades para influir en el beneficio de la empresa.

2. **Interacción entre MTenure y MgrSkill:** Esta variable refleja cómo la antigüedad en el puesto de los gerentes interactúa con sus habilidades. Observar cómo la experiencia acumulada de los gerentes se relaciona con sus habilidades puede proporcionar información valiosa sobre su impacto en el rendimiento financiero de la empresa.

La inclusión de estas nuevas variables de interacción en nuestro modelo de regresión lineal nos permitirá examinar cómo la experiencia y las habilidades conjuntas del personal y los gerentes afectan al beneficio de la empresa. Su significancia estadística y su impacto en la predicción del beneficio nos ayudarán a obtener una comprensión más completa de los factores que influyen en el éxito financiero de la empresa.

```
## [1] 63 11
```

```
## [1] 12 11
```

```
##
```

```
##
```

Métrica	Valor
RMSE (Error Cuadrático Medio Raíz)	0.3179117
R <sup>2</sup> (Coeficiente de Determinación)	0.9243495

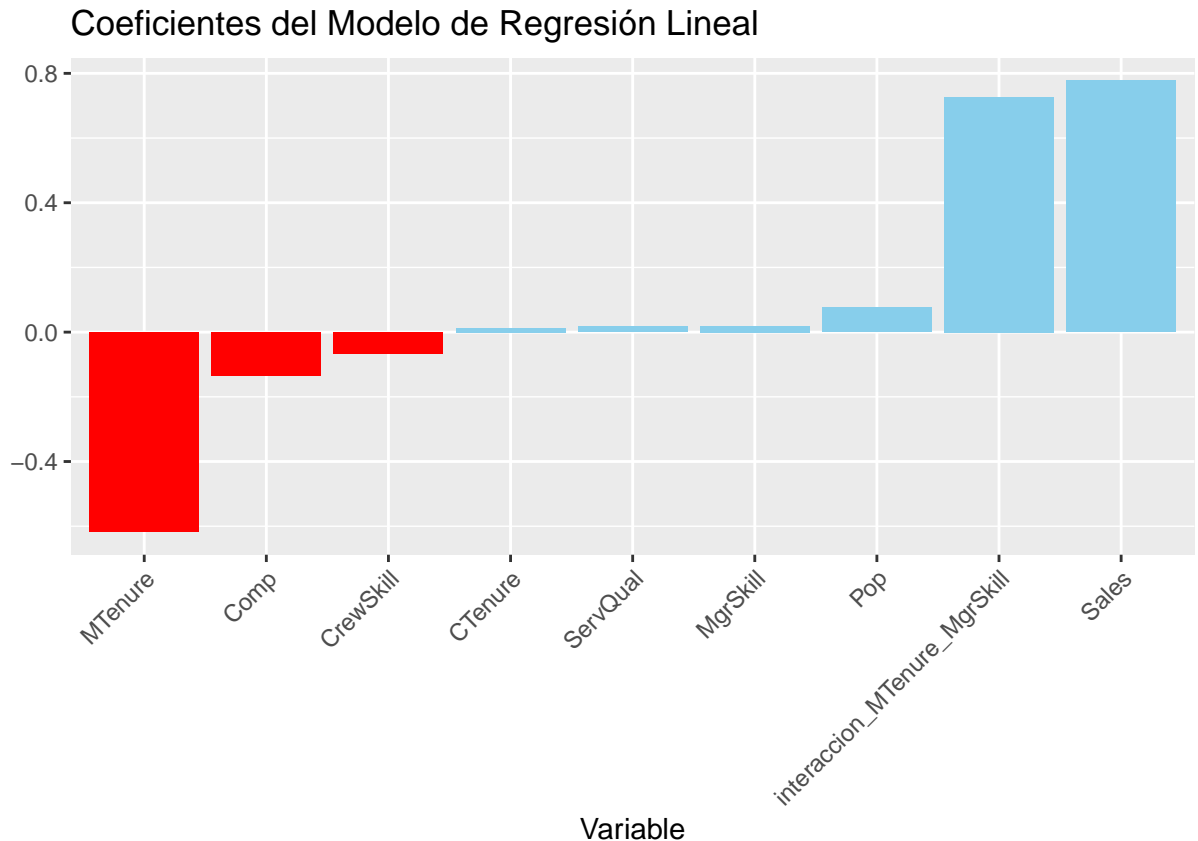
```
##
```

```
##
```

```
## Table: Coeficientes del Modelo de Regresión
```

```
##
```

	Variable	Coeficiente
(Intercept)	(Intercept)	-0.0136145
Sales	Sales	0.7784563
MTenure	MTenure	-0.6182811
CTenure	CTenure	0.0125040
Pop	Pop	0.0761084
Comp	Comp	-0.1346759
CrewSkill	CrewSkill	-0.0650698
MgrSkill	MgrSkill	0.0189681
ServQual	ServQual	0.0172254
interaccion_MTenure_MgrSkill	interaccion_MTenure_MgrSkill	0.7272289



Para este modelo, se ha incorporado una nueva variable que es el producto de MTenure (Antigüedad en el puesto del gerente) y MgrSkill (habilidad de gestión). Es notable observar que esta nueva variable resulta altamente significativa en nuestro análisis. Esto sugiere que la combinación de la antigüedad en el puesto del gerente y su habilidad de gestión puede desempeñar un papel crucial en el resultado de nuestras ganancias. Este hallazgo amplía nuestra comprensión de los factores que influyen en el rendimiento de las tiendas, y subraya la importancia de considerar interacciones entre variables en futuros análisis.

#### Cálculo de bonificación para un gerente que presenta los siguientes datos en su tienda

	store	Sales	Profit	MTenure	CTenure	Pop	Comp	CrewSkill	MgrSkill	ServQual
1	9	2113089	474725	108.9935	6.061602	26519	2.63763	3.22	3.583333	100
2	9	2747015.7	901977.5	130.7922	7.2739224	26519	2.63763	3.22	3.583333	100

```
##          2
## 49349.73
```

#### Resultado de la aplicación del modelo:

La estimación de las ganancias en la tienda 9 bajo estos nuevos parámetros es de 49.349,73 y el promedio de los meses de servicios del gerente aumentó en 6.6 meses en promedio, con este valor puede estimarse una bonificación al gerente.

RPubs: <https://rpubs.com/jaquino61/1161054>